DOCUMENT RESUME

ED 341 430 JC 920 091

AUTHOR Boylan, Hunter R., Ed.; Kerstiens, Gene, Ed.
TITLE Assessment.
INSTITUTION Appalachian State Univ., Boone, NC. Center for Developmental Education.
PUB DATE 89
NOTE 26p.
AVAILABLE FROM Center for Developmental Education, Appalachian State University, Boone, NC 28608.
PUB TYPE Collected Works - Serials (022) -- Information Analyses (070) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120)
JOURNAL CIT Research in Developmental Education; Pilot n1, v1 n5, v3 n1, v6 n5 1983,1985,1989

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; *Basic Skills; Community Colleges; Educational Testing; Higher Education; High Risk Students; Postsecondary Education; *Remedial Instruction; School Surveys; *Standardized Tests; *Student Evaluation; *Student Placement; Testing Programs; Test Interpretation; Test Reviews; Test Use; *Test Validity

ABSTRACT

These four serial issues examine the effectiveness and appropriateness of a variety of assessment tests as well as their relationship to developmental education. Included are reviews of the following tests: (1) the Comparative Guidance and Placement Program, a self-scoring test of English and mathematics; (2) the Stanford Achievement Test, an advanced battery of tests of vocabulary, reading comprehension, mathematical concepts, computation and application, science, and social sciences; (3) the Comprehensive Test of Basic Skills, a test to assess proficiency in reading, language skill, arithmetic, science, social science, and study skills; (4) the Nelson-Denny Reading Test, which measures reading comprehension, vocabulary, and reading rate; (5) the California Achievement Test, measuring reading levels from grades 1 through 12 by testing reading comprehension and vocabulary; (6) the Sequential Test of Educational Progress, used to assess "higher order" intellectual skills related to reading, such as comprehension, inference, analysis, and translation; (7) the Canfield Learning Styles Inventory, a test to assess the affective dimensions of learning such as student preferences for conditions of learning, the content of learning, and the mode of learning as well as student expectations; and (8) the Kolb Learning Styles Inventory, a test to measure the cognitive dimensions of learning styles, such as students' use of reflective observation, abstract conceptualization, and active experimentation. In addition, "Assessing Assessment," by Dennis Gabriel, is presented, which discusses the results of four surveys of educational institutions on assessment strategies and test use, and which calls for more widespread use of mandatory testing and placement of underprepared students in basic skills courses. (PAA)

SELECTED ISSUES OF: "Research in Developmental Education"

REVIEW OF

RESEARCH in Developmental Education

Published by Appalachian State University

ED341430

ON: Assessment. Pilot Issue #1, 1983
Volume 1 Issue 5, 1983
Volume 3 Issue 1, 1985
Volume 6 Issue 5, 1989

JC920091

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Garbriel

H. R. Boylan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Research in Developmental Education

"Pilot Issue #1, 1983"

About "RDE".....

Believe it or not, practically the entire field of developmental education is based upon research. Research activities first suggested that the "open door" was in danger of becoming a "revolving door" without some sort of developmental intervention for underprepared students. Once these interventions were adopted on college and university campuses, research formed a basis for the practice of individualized instruction, mastery-learning, self-concept development, diagnosis, prescription, and a host of other activities carried out through developmental programs.

Yet the typical developmental educator does not have the time to review the existing research. While a wealth of research information relating to developmental education is available, it is often too difficult to find, or having been found, too arcane to interpret.

It is this situation that makes "Research in Developmental Education" (RDE) such a potentially valuable tool for practitioners in the field. RDE is designed to review current research in areas relating to the practice of developmental education. Furthermore, it will attempt to interpret this research in terms of its applicability to developmental education programs.

NEW! NEW! NEW!

A New Publication From
THE
CENTER FOR DEVELOPMENTAL EDUCATION

***

Featuring

***

PLACEMENT TESTS:

Evaluation & Comments

*** PAGE 2! ***

and

*** RESOURCERY ***
Your Guide to More
PAGE 3!

***

We plan to publish RDE at least five times each year. Each issue will be devoted to research on a particular topic of interest to developmental educators. Each issue will include a review of relevant research, a summary of research findings, a list of suggested applications for these findings, and suggested resources for learning more about the topic under consideration. It is hoped that such a publication will offer a viable link between research and practice.

Of course, like any other publication, RDE will take time and money to produce. Although it is provided as a service to the field of developmental education, we must pay for reproduction, printing, typing, and mailing. As a result, RDE is being offered on a subscription basis. Our initial subscription rate will be $9.50 per year. This includes a volume of 5 issues appearing in September, November, January, March, and May. We will offer RDE during the 1983 calendar year and then review the situation in January of 1984. If we can meet our costs as of that time, we will continue to publish and, we hope, expand the newsletter. If we cannot meet our costs, we will discontinue its publication.

Needless to say, your cooperation and support is necessary if RDE is to become a long-term reality instead of a short-lived good idea. What can you do to help?... First of all, we need your subscription to RDE.

We also need your advice and feedback. What topics would be of interest to developmental educators? How can we improve the content or the lay-out of RDE? What sort of information should we provide regarding resources? How can we improve dissemination of RDE?

If you have any comments or suggestions in any of these areas, please let us know. We will look forward to hearing from you (and to receiving your subscription request for the next five issues of RESEARCH IN DEVELOPMENTAL EDUCATION).

THANK YOU!

Hunter R Boylan

1

Placement Tests Evaluation & Comments

One of the activities most frequently encountered in developmental programs is testing designed to obtain information for placement. Cross (1976) found that over 80% of the programs she surveyed in 1974 utilized some sort of standardized tests to place students in developmental courses. Roueche and Snow found similar results in their survey of developmental programs (1977).

A 1979 survey conducted by the Center for Developmental Education suggested that two instruments used most frequently for placement purposes in developmental programs were the COMPARATIVE GUIDANCE AND PLACEMENT PROGRAM (CGP) and the STANFORD ACHIEVEMENT TEST. In recent years, many developmental programs have also adopted the COMPREHENSIVE TEST OF BASIC SKILLS (CTBS) for placement purposes.

Since these three tests seem to be the most widely-used comprehensive placement instruments in developmental programs, it is appropriate for the inaugural issue of RESEARCH IN DEVELOPMENTAL EDUCATION to review their advantages and disadvantages as noted in current research. Before looking at the individual tests, however, a word advice regarding the use of placement instruments may be in order.

As Glaser (1971, p. 8) notes, norm-referenced placement instruments provide information regarding students' "...relative standing along a continuum of attainment... they tell that one student is more or less proficient than another but not how proficient either of them is with respect to the subject matter tasks involved." As a result, such placement instruments are measures only of what a student can do at the time of testing. They do not measure student potential for learning nor do they provide precise measurement of student deficiencies. Placement tests, therefore, are best used in sorting students into broad categories. They are less suitable for predicting how well a given student might perform or for diagnosing student weaknesses.

Assuming that placement instruments are to be used in sorting students for advisement and scheduling purposes, just how good are the tests most commonly used in developmental programs? The answer to this question can be found in the psychometric literature.

THE CGP

The COMPARATIVE GUIDANCE & PLACEMENT PROGRAM, published by the College Entrance Examination Board for the Educational Testing Service, is a self-scoring test of English and mathematics. The test includes sections on reading and written English expression and in mathematics computation, applied arithmetic, elementary and intermediate algebra. Of all the tests reviewed, the CGP has the shortest administration time -- a maximum of 105 minutes for all six sections.

In reviewing the CGP, Hambleton suggests that the instrument's reliability has yet to be firmly established (1978). Reliability data provided in the CGP Technical Manual (1975) is insufficient to establish the degree to which student scores on the instrument will vary significantly from one administration to another. This potential weakness of the CGP is also noted by Zytowski (1974).

The validity of the CGP has also been questioned by reviewers. Hastings (1978) has suggested that the content validity of the instrument is weak and as a result it is difficult to determine exactly what the test measures. The CGP is also somewhat lacking in predictive validity. Correlations between CGP scores and later student performance in the skill areas measured by the test seldom exceed .40 -- a correlation coefficient of only slight significance (Zytowski, 1974).

Although the CGP may have some limitations to its validity, these may be overcome by developing local norms as recommended by Maxwell (1979). Such norms, when matched against later student performance, can increase the validity of decisions based on CGP scores.

When local norms are used in making placement decisions based on the CGP, the instrument can serve as a valuable placement tool. As Maxwell points out, the CGP has the advantage of saving "...both student and advisor time because the results are immediately available to use in planning a program and scheduling a student into appropriate skills courses" (1979, p. 44).

THE STANFORD TEST

The Stanford Achievement test, published by the Psychological Corporation, was designed originally for assessment of elementary stu-

Resourcery

Resource Books

Anastasi, Anne (Editor), Testing Problems in Perspective, Washington, D.C.: American Council on Education, 1966.

Anderson, S. B., Ball, S., Murphy, R. T., et. al., Encyclopedia of Educational Evaluation, San Francisco: Jossey-Bass, 1975.

Beatty, W. (Editor), Improving Educational Assessment, Washington, D. C.: Association for Supervision and Curriculum Development, 1969.

Brown, F. G., Principles of Educational and Psychological Testing, New York: Holt, Rinehart, and Winston, 1976.

Buros, O. K. (Editor), Tests in Print: II, Highland Park, NJ: Gryphon Press, 1974.

Buros, O. K. (Editor), Eighth Mental Measurements Yearbook, Highland Park, NJ: Gryphon Press, 1978.

Gronlund, N. E., Measurement and Evaluation in Teaching, New York: MacMillan & Company, 1967.

Lien, A. J., Measurement and Evaluation of Learning, Dubuque, IA: Wm. C. Brown, 1976.

Popham, W. J., Educational Evaluation, Englewood Cliffs, NJ: Prentice-Hall, 1975.

Stodola, Q. S. and Stordahl, K., Basic Educational Tests and Measurement, Chicago: Science Research Associates, 1967.

Thorndike, R. L. and Hagen, E. P., Measurement and Evaluation of Psychology and Education, New York: John Wiley & Sons, 1977.

Wallace, G. and Larsen, S. C., Educational Assessment of Learning Problems: Testing for Teaching, Boston, Allyn and Bacon, Inc., 1978.

Publishers

AMERICAN COLLEGE TESTING PROGRAM (ACT), P. O. Box 168, Iowa City, IA, 52240 - (319) 351-4470

CTB/McGRAW-HILL, Del Monte Research Park, Monteray, CA, 93940 - (408) 373-2932

EDUCATIONAL TESTING SERVICE (ETS), Princeton, NJ, 08540 - (609) 921-9000

HARCOURT BRACE JANOVICH, INC., 757 Third Avenue, New York, NY 10017 - (212) 572-5000

HOUGHTON-MIFFLIN COMPANY, Pennington-Hopewell Road, Hopewell, NJ 08525 - (609) 466-1950

JASTAK ASSOCIATES, INC., 1526 Gilpin Avenue, Wilmington, DE 19806 - (302) 652-4990

THE PSYCHOLOGICAL CORPORATION, 304 East 45th Street, New York, NY 10017 - (212) 679-7070

SCIENCE RESEARCH ASSOCIATES, INC., 259 East Erie Street, Chicago, IL 60611 (312) 266-5000

---

cont. from p. 6 -

cision of measurement is important.

The best placement testing programs are likely to be those which strike a thoughtful balance between various testing considerations. Ease of administration, scoring, interpretation, advising, placement procedures, and level of precision required are all important considerations in designing a testing program. It is important to remember, however, that placement instruments are best used when making broad categories of decisions regarding enrollment in a particular course or curriculum. When used in this manner, placement tests can be powerful tools for promoting academic development. When used for other purposes, they may, at best, be of marginal value and, at worst, be destructive to to students and instructors alike.

REFERENCES

Ahmann, J. S., in Oscar Buros Seventh Mental Measurement Yearbook, Highland Park, NJ: Gryphon Press, 1972, p. 38.

Biemiller, L., "Microcomputers Expected to Bring Radical Changes in Testing," Notes from the National Testing Network in Writing, Vol. 1, No. 3, January, 1983, p. 15.

Comprehensive Guidance and Placement Program Technical Manual, Monteray, CA: CTB/McGraw-Hill, 1975.

Ebel, R. L., in Oscar Buros Eighth Mental Measurements Yearbook, Highland Park, NJ: Gryphon Press, 1978, p. 56.

Cross, K. P., Accent on Learning. San Francisco: Jossey-Bass, 1976.

Findley, W. G., in Oscar Buros, Eighth Mental Measurements Yearbook, Highland Park, NJ: Gryphon Press, pp. 40-43.

Glaser, R., "Instructional Technology and the Measurement of Learning Outcomes: Some Questions," in W. James Popham, Criterion-Referenced Measurement, Englewood Cliffs, NJ: Educational Technology Publications, 1971, p. 8.

Hambleton, R. K., in Oscar Buros, Eighth Mental Measurements Yearbook, Highland Park: NJ, Gryphon Press, 1978, pp. 17-18.

Hastings, T. J., in Oscar Buros, Eighth Mental Measurements Yearbook, Highland Park, NJ: Gryphon Press, 1978, pp. 18-19.

Lehmann, I. J., "Review of the Stanford Achievement Test," Journal of Educational Measurement, Vol. 12, No. 4, Winter, 1975, pp. 297-306.

Maxwell, M., Improving Student Learning Skills, San Francisco: Jossey-Bass, 1979.

Roueche, J. E. and Snow, J., Overcoming Learning Problems, San Francisco, Jossey-Bass, 1977.

Technical Bulletin: Form S of the CTBS, Monteray, CA: CTB/McGraw-Hill, 1976.

Zytowski, D. G., "Review of the Comparative Guidance and Placement Program," Journal of Counseling Psychology, Vol. 21, No. 3, May, 1974, pp. 254-258.

dents. The recent addition of an advanced battery providing readings at the 7.0 to the 9.9 grade level makes the instrument acceptable for use in basic skill assessment of college students.

The advanced battery includes tests of vocabulary, reading comprehension, mathematics concepts, computation and application, science and social science. The Stanford test takes considerably more time to administer than the CGP. As much as 315 minutes may be required to administer all sections of the advanced battery.

Available evidence suggests that the Stanford is a highly reliable instrument. Technical data prepared by the publishers and reviewed by Ebel (1978) indicates that the reliability coefficients for most sections of the test are at or above the .90 level. This is a very high index of reliability and it suggests that little fluctuation from one administration of the test to another may be expected in student scores.

Insofar as validity is concerned, the Stanford has received mixed reviews although it is generally considered sound. Lehmann (1975) has argued that insufficient data has been provided by the publishers to determine the instrument's content validity. This criticism is also noted by Ebel (1978). On the other hand, Passow (1978) points out that the test's content validity is established by the fact that classroom instructors were consulted at every step in the test's development and that they assisted in designing questions and reviewing content. This, alone, should serve to support the content validity of the Stanford.

The Stanford also appears to possess a high degree of concurrent validity as it correlates well with other achievement measures. The various sections of the test also correlate well with one another. In fact, Kasdon (1974) points out that the science and social science components of the test are so highly inter-correlated with other components that it may not be necessary to administer the entire battery in order to obtain accurate information.

THE CTBS

The COMPREHENSIVE TEST OF BASIC SKILLS (CTBS) is published by McGraw-Hill. It is designed to assess student proficiency in the areas of reading, language skill, arithmetic, and study skills. The newer version of the test (forms T, U, and V) also provides a measure of science and social science skills. Level IV of the CTBS measures skills in the grade ranges of 8.5 to 12.9 making it quite appropriate to college students.

Like the Stanford test, the CTBS possesses a high degree of reliability. Ahmann (1972) reports that the reliability coefficients for subsections of the test range from .85 to .95 thus making it a highly reliable instrument.

The CTBS also appears to have adequate validity -- particularly content validity. The CTBS Manual (1976) provides data supporting the content validity and concurrent validity which suggest that the instrument is quite sound in these areas. In his review of the CTBS, Ahmann agrees that the content validity of the test is adequate. He also notes that the CTBS is strongly correlated with the CALIFORNIA BASIC SKILLS TEST. In fact, he raises the question as to why it was necessary to produce the CTBS at all since it is so highly correlated with the California test (also published by McGraw-Hill). Putting the question of redundancy aside, however, the high inter-correlation between these two instruments indicates that the CTBS possesses strong concurrent validity.

One of the advantages of the CTBS is that, of all the tests reviewed here, the designers of the CTBS have done the best job of controlling for cultural bias. According to Findley (1978), bias was controlled for by: 1) having all test items reviewed by qualified editors to eliminate any items that might contribute to bias and, 2) eliminating all items on which minorities scored less well than the standardized sample but which were not positively correlated with later performance. These procedures make the CTBS one of the least biased placement instruments on the market.

In spite of its many desireable qualities, the CTBS does have one major disadvantage. It takes a rather substantial amount of time to administer. If the complete test battery is administered, maximum time required may be as much as 335 minutes. The time required for administration, however, may be reduced by eliminating some of the less relevant sub-tests such as science and social science.

SUMMARY & RECOMMENDATIONS

All of the tests reviewed here have strengths and weaknesses as placement measures. The CTBS appears to have a

slightly greater degree of content validity than the Stanford and a considerably greater degree than the CGP. On the other hand, it is also the longest of the three instruments -- taking more than three times as long to administer than the CGP.

Both the CTBS and the Stanford are highly reliable instruments that will give consistent scores from one administration to another. The CGP appears to be somewhat less reliable than the other two instruments.

The CTBS has the best controls for bias of any of the tests reviewed. However, the CGP is normed with a group containing a very high percentage of minorities. The norming process of the CGP may, therefore, serve to reduce the possibility of cultural bias in the instrument.

All of the tests reviewed appear to lack predictive validity. In other words, they may not be good predictors of actual student performance in a given class. This, however, is not surprising. After all, the content of the instruments is fixed while the content of courses is likely to vary. For this reason, the development of local norms and the correlation of these norms with later student performance is strongly recommended.

In spite of the amount of research available to judge the quality of various tests, the question "which test is best for placement in developmental courses?" cannot be answered simply. The "best" placement instrument is the one which best meets local needs. If ease of administration and scoring is of primary importance, then the CGP may be appropriate in spite of its limitations. This is particularly true if local norms are developed to assist in placement decisions and if initial decisions can be adjusted easily.

If precision in measurement is a primary consideration, however, the CTBS or the Stanford are likely to provide more precise information. The amount of time required to administer these instruments may be reduced, if necessary, by administering only certain components of the tests. The reading, vocabulary, and mathematics sections of both the CTBS and the Stanford are both reliable and valid. They also correlate well with other components. Administration time might be reduced without compromising the quality of information provided by using only these test components.

As a general rule, the more difficult it is to adjust initial placement decisions, the more important it will be to obtain precise information as a result of placement testing. In situations where students may move rather easily from a developmental course to a more advanced course in the first few weeks (or vice versa), then incorrect decisions regarding placement may be less damaging. In situations where initial placement decisions are difficult to reverse, then pre-

continued, page 4

****

SUBSCRIBE NOW!

To receive RESEARCH IN DEVELOPMENTAL EDUCATION, send $9.50 to:

Managing Editor
Center for Developmental Education
Appalachian State Univ.
Boone, NC 28608

****

Research in Developmental Education is published five times per academic year.
Editor: Hunter R. Boylan
Managing Editor: Doree N. Pitkin
Consulting Editor: Milton "Bunk" Spann

Manuscripts, news items, and abstracts are accepted by the Editor, RiDE, Center for Developmental Education, Appalachian State University, Boone, NC 28608.

Subscriptions are $9.50 per year. North Carolina residents add 38 cents sales tax; subscribers in foreign countries and $1.50/year shipping and pay by bank draft. Send subscriptions to Managing Editor, RiDE, at the same address or call (704)262-3057.

© 1983

Center for Developmental Education
Appalachian State University
Boone, North Carolina 28608

Research in Developmental Education

Volume 1
Issue 5

A Review of Diagnostic Reading Tests

By: Hunter R. Boylan

Practically all developmental and learning assistance programs provide some sort of reading tutoring or instruction. In the majority of programs, this training in reading is predicated on the assessment of students' existing reading skills. It is from this assessment base that most reading instruction or tutoring proceeds. Consequently it is important for those who work in developmental programs to know what options are available to them for assessment of student reading skills. This issue of RiDE is, therefore, devoted to a review of the reading assessment instruments most commonly used in developmental programs.

BACKGROUND AND METHOD

To date, no national survey data is available to determine the types of diagnostic reading tests used most commonly in developmental programs. However, several regional studies have been undertaken. The Arkansas Consortium for Developmental Education surveyed reading instruments used by practitioners in the State of Arkansas in 1978 (ACDE Newsletter). The Center for Developmental Education surveyed the diagnostic activities of North Carolina colleges participating in a regional consortium in 1979 (Center for Developmental Education, 1980). Perhaps the most important current study was completed by the Washington Association for Developmental Education under a FIPSE grant in 1982. This survey reported data from 28 community colleges in the State of Washington and provided information on user reactions and recommendations for the use of various tests (WADE, 1983)

****
IN THIS ISSUE
****

Dr. Hunter Boylan
Reviews
Diagnostic Reading
Tests

****

Nelson-Denny
California Achievement
CGP
Stanford Achievement
STEP

Published by
Appalachian State University

While these surveys are regional, the reported use of reading assessment tests was fairly consistent in each survey. It seems reasonable to conclude, therefore, that they are somewhat representative of developmental and learning assistance programs across the country. While the degree of this representation is unknown, the information from these studies at least provides a baseline from which to make informed judgements about reading test utilization.

Based on the available data, an informed judgement would suggest that only a handful of reading instruments are used with any degree of regularity in developmental and learning assistance programs. The surveys reviewed included a total of 68 institutions. Among these institutions, only 5 reading tests were used with regularity.

These include the Nelson-Denny Reading Test and the reading sections of the California Achievement Test, the Comprehensive Guidance and Placement Test, the Stanford Achievement Test and the Sequential Test of Educational Progress.

The Nelson-Denny Reading Test was the most widely used assessment instrument with the reading section of the Comprehensive Guidance and Placement Test being the next most widely-used. The California Achievement Test, the Stanford Achievement Test and the Sequential Test of Educational Progress were all used with about equal regularity.

A cautionary note should be added here since the vast majority of institutions included in the surveys reviewed were community and technical colleges. Of the

institutions responding to various surveys, 58 were two-year schools and only 11 were four-year schools. As a result, the comments made here are far more applicable to two-year institutions.

After identifying the five most frequently used reading diagnostic tests, user responses were analyzed from the reports of WADE and the Center for Developmental Education (the Arkansas report did not include user responses). Buros' MENTAL MEASUREMENTS YEARBOOK (Eighth Edition, 1978) was then consulted to obtain technical information on these tests.

THE CALIFORNIA ACHIEVEMENT TEST

The California Achievement Test, published by CTB/McGraw-Hill, includes a reading component measuring vocabulary and comprehension. It measures reading levels from grades 1 through 12 and results are reported in raw scores, percentiles, and stanines. The reading comprehension section of the test is considered by reviewers to be a sound measure of comprehension The vocabulary section is, however, somewhat less precise. Since this section includes only 40 words, missing one or two items may make a big difference in placement results.

The items included in the California Achievement Test are drawn from an extensive review of recommended curiculum objectives from several state boards of education.

The California Achievement Test is essentially an achievement test for the elementary and secondary levels. Its relevance to college-level placement is dependent upon the degree to which a given institution's curriculum is keyed to the objectives of the public school curriculum.

A potential problem with the California Achievement Test is that some reviewers consider it to be sexually biased. A few of the terms used may be interpreted differently by women and answered incorrectly as a result (Lombard, 1978). This potential difference does not seem to be reflected in normative data for men and women but it may make a slight, statistically insignificant, difference in placement.

Reliability

The reliability of the comprehension section of this instrument is strong. The publishers cite a wide variety of studies establishing the internal consistency of the measure with correlation coefficients in the .65 to .80 range. The vocabulary section appears to be slightly less reliable but still strong.

Validity

The publishers claim that it is only necessary to establish content validity in an instrument that is, essentially, a national achievement test. Content validity has been established by matching items against recommended curriculum objectives. The content of test items does appear to match these objectives fairly well and the test does appear to have strong content validity.

User Comments

Most users seem to regard the California Achievement Test as valid when used for general placement purposes. It may be used in making initial placement decisions but it does not provide specific information for diagnostic purposes. User comments also suggest that the reading passages may not be appropriate for older adults.

THE COMPREHENSIVE GUIDANCE AND PLACEMENT PROGRAM

The Comprehensive Guidance and Placement Program, better known as the CGP, is a self-scoring placement battery produced for the College Entrance Examination Board by ETS. Among community colleges, it is probably the most widely-used general placement battery. The reading section of the battery is relatively brief and is primarily a measure of comprehension.

While the CGP has the advantage of being quick and easy to administer, it is not a particularly precise instrument. The precision of the instrument may be improved, however, by gathering local data for normative purposes. Local norms may then be developed to establish cut-off points for placement in the college curriculum (Maxwell, 1978).

The CGP was originally developed for

BEST COPY AVAILABLE

community college placement. The basic construct of the CGP is that there is an identifiable body of skills associated with college-level work and that these skills can be identified, and items canbe developed that will match these skills.

Since the level of academic skill required for success varies dramatically from one college to another, it is doubtful that the CGP can accomplish this purpose without local normative data. With such data the CGP is probably a reasonably valid general placement instrument. It also has the advantage of being keyed to what the publishers consider to be college-level academic competencies rather than to high school-level competencies.

Reliability

The reliability of the reading section of the CGP is not particularly strong. Correlation coefficients for internal consistency range from .35 to .60 and the reading section's correlation with other sections of the test is also low.

Validity

The publishers attempted to establish content validity by having test items reviewed by community college faculty. The purpose of this review was to determine whether or not the items were consistent with the community college curriculum. The CGP is one of the few instruments in which actual community college faculty were involved in validation of test items.

The publishers have attempted to establish the predictive validity of the instrument by correlating CGP results with student grades. The correlations were extremely low (ranging from .30 to .40). The predictive validity of theCGP may, however, be improved by using local norms for prediction purposes (Buros, 1978).

User Comments

Users tend to like the CGP because it is quick and easy to score. It does allow for faster testing of a larger number of students than most other instruments. Furthermore, since the test may be self-scored, there is no need to wait for computer processing of test results. Students can obtain their test scores almost immediately after taking the test. This immediate feedback is certainly an advantage.

Most users are aware of the instrument's shortcomings as a placement device. Users consistently suggest that it be used only as a very generalized placement measure and that placement decisions be made on the basis of local norms.

THE NELSON-DENNY READING TEST

The Nelson-Denny Reading Test is the most widely-used instrument among those responding to the surveys reviewed. Maxwell (1978) suggests that it may also be the instrument most widely used by 4-year colleges and universities for initial placement purposes. The publishers claim that the test is designed to provide a trustworthy ranking of student ability in the areas of reading comprehension, vocabulary development, and reading rate. The construct used in designing the test was that normative data from college students will yield a valid measure of how well such students ought to be able to read at various college grade-level equivalents. Grade level reports, therefore, are based on normative data rather than any particular curriculum objectives. Reviewers generally suggest that this construct is valid and that the test does accomplish the purposes for, which it was designed (Forsyth, 1978 and Cummins, 1981).

The Nelson Denny is well regarded for college placement purposes because it is designed for college students and it is relatively quick and easy to administer. the most recent form of the test (forms E and F) also allows self-scoring.

The instrument measures reading skills at the level of 9th through the 16th grades. However, owing to limitations in the norming sample, scores from grades 13 to 16 are less reliable than scores for the lower levels. Grade level placement is also questionable because only a few items will make a difference of one or more grade levels. Raw scores, stanines and percentiles are also reported on the Nelson-Denny and these may be better indicators of students' actual performance.

Reliability

The Nelson-Denny Reading Test tends to have fairly high reliability for its vocabulary and comprehension sections. The range of reliability scores reported for these sections are .82 to .91 and .68 to reading rate assessment is considerably lower, ranging from .54 to .66.

Validity

The Nelson Denny Reading Test has been validated primarily through norm-referencing. The publishers provide a variety of of technical data from studies conducted with appropriate samples of college students taken at various points in the academic year. The instrument has also been reviewed for content validity by curriculum experts. The instrument appears, on the basis of these tests, to be quite valid when used as recommended.

User Comments

In general, users are quite satisfied with the quality of information yielded by the Nelson-Denny as long as the results are used for general placement purposes. While the Nelson-Denny is often used as a diagnostic instrument, it does not provide sufficiently precise data to diagnose specific strengths and weaknesses. Users also caution against relying on alternative forms of the Nelson-Denny for pre and post-test investigation of reading gains. Apparently, pre- and post-test results are heavilyinfluencedbytestfamiliarity (Cummins, 1981).

THE SEQUENTIAL TEST OF EDUCATIONAL PROGRESS

The Sequential Test of Educational Progress (STEP) is produced by the Educational Testing Service and published by the Addison-Wesley Company. Unlike other reading instruments reviewed here, the STEP is designed to assess so-called "higher order" intellectual skills such as comprehension, inference, analysis, and translation (Wardrop, 1978). Items in the reading section of the test are selected on the basis of how well they measure these constructs. Because of the importance of such skills in the college-level curriculum, the STEP is generally regarded as a good placement instrument for college students althoughitwasoriginally based on elementary and secondary students.

The STEP reading section includes 60 items, 30 to test vocabulary and 30 to test comprehension. Unfortunately, only a combined scorefrom both sectionsis provided. As a result, the STEP is difficult to use for diagnostic purposes (Johnson, 1978). This is unfortunate since more specific information on higher order intellectual skill strengths and weaknesses would be quite valuable to developmental educators.

The STEP reports scores in grade level equivalents, ranging from grades 4 to 14. It also reports stanine and percentile scores. Johnson (1978) notes that, for reasons as yet unknown, women tend to outperform men on the reading section of the STEP by as much as one score point.

Reliability

Reliability tests for the STEP were conducted using measures of internal consistency and analysis of test-retest reliability using alternative test forms. In both areas, the STEP demonstrated a high degree of reliability. The most recently reported ranges for reliability were between .76 and .93 (Johnson, 1978).

Validity

The test developers established content validity through expert review of the items to insure that they were related to the higher order skills being assessed. No data has yet been provided to establish the construct validity of the instrument.

User Comments

Because the vocabulary and comprehension scores are combined in the STEP, users are frequently disatisfied with the utility of results. Furthermore, since the test is based on a set of theoretical constructs rather than specific curriculum objectives, it haslittle predictive validity. The STEP may be more useful as a measure of intellectual skills than as a placement or diagnostic instrument.

THE STANFORD ACHIEVEMENT TEST

The Stanford Achievement Test, published by the Psychological Corporation,

is one of the earliest and most reliable reading assessment instruments in existence. The test was originally designed for the assessment of elementary school students and, for many years, the cut-off point for measurement was at the 6th grade level. More recent editions of the test measure skills up to grade level 9.9. This comparatively low cut-off range makes it appropriate only for basic skills students at the college level.

The Stanford Achievement Test measures vocabulary, comprehension, and word attack skills. Test items used to measure these skills are based on analysis of school textbooks, analysis of curriculum objectives, and expert review. Since the test is keyed to the public school curriculum, its relevance to college level placement is questionable. The Stanford Achievement Test is much like the California Achievement Testin this regard. Glass suggests that there is really little difference between the two instruments (1978).

The Stanford Achievement Test provides grade equivalent scores, percentile ranks, and stanines. These scores are normed on a substantial population including 275,000 school children from 43 different states. As a result, the normative data for this instrument is probably the strongest of any of the reading tests reviewed here.

Reliability

Since the Stanford Achievement Test has been in existence for so many years and has been so widely-used in the public schools, reliability has been continually analyzed and improved. It is probably one of the most reliable instruments on the market with most reliability coefficients ranging from the .85 to .95.

Validity

Like the California Achievement Test, the Stanford is designed as a national measure of school achievement. As a result, the publishers claim that the only appropriate form of validity for assessing the instrument is content validity. Over the years, a variety of testing experts have analyzed this instrument and carefully verified this content validity.

User Comments

Most users consider the Stanford Achievement Test to be useful for general placement of underprepared students. This is particularly true for those who are recent high school graduates. Because of its strong elementary school content, however, its use with adults may be quesionable. Similarly, most users indicate that the instrument has relatively little predictive validity.

SUMMARY

Of the tests reviewed here, the Nelson-Denny Reading Test appears to be most applicable to college level placement. It is valid, reliable, and easy to use. It is normed specifically for use with college level students and it reports vocabulary, comprehension, and reading rate scores.

The Nelson-Denny does, however, have its limitations. Its measurement at the upper grade levels is less reliable than at the lower grade levels. This is further complicated by the fact that only a few items will make a difference between placement in one grade level or another. Also, the Nelson-Denny does not provide sufficient information to make it usable for specific diagnosis of reading problems. It appears to be best used as a generalized placement instrument.

There may, however, be excellent reasons for using a placement instrument other than the Nelson-Denny. The Nelson-Denny is specifically a reading test. It does not provide placement data in other subject areas. Instruments such as the CGP or the California Achievement Test do provide a complete battery of tests for placement purposes. Also, the Nelson-Denny may be too difficult for underprepared students. Other instruments provide much broader placement ranges that may be more suitable to severely underprepared students.

Perhaps the most important point to be made here is that none of the tests reviewed are particularly useful as diagnostic instruments. They are used best as pre-screening devices to give practitioners general information on where to begi in working with developmental students.

REFERENCES

Arkansas Consortium for Developmental Education, ACDE NEWSLETTER, Spring, 1979.

Center for Developmental Education, "Instruments for Measuring Cognitive and Affective Development Among Students," Appalachian State University, Boone, NC, 1980.

Buros, Oscar, SEVENTH MENTAL MEASUREMENTS YEARBOOK, Highland Park, NJ: Gryphon Press, 1978.

Cummins, R. Porter, "The Nelson-Denny Reading Test (Forms E and F)," JOURNAL OF READING, October, 1981, pp. 54-59.

Glass, Gene V. in Buros SEVENTH MENTAL MEASUREMENTS YEARBOOK, Highland Park, NJ: Gryphon Press, 1978, pp. 1207 - 1209.

Forsyth, Robert, in Buros SEVENTH MENTAL MEASUREMENTS YEARBOOK, Highland Park, NJ: Gryphon Press, 1978, pp. 1209-1211.

Johnson, Richard in Buros, SEVENTH MENTAL MEASUREMENTS YEARBOOK, Highland Park, NJ: Gryphon Press, 1978, pp. 1219-1220.

Lombard, Richard, in Buros, SEVENTH MENTAL MEASUREMENTS YEARBOOK, Highland Park, NJ: Gryphon Press, 1978, p. 1205.

Maxwell, Martha, IMPROVING STUDENT LEARNING SKILLS, San Francisco: Jossey-Bass, 1978.

Wardrop, James, in Buros, SEVENTH MENTAL MEASUREMENTS YEARBOOK, Highland Park, NJ: Gryphon Press, 1978, p. 1222.

SUBSCRIBE TO RiDE TODAY!

Send your name and mailing address along with a check or purchase order for $9.50 to:

MANAGING EDITOR, RiDE
CENTER FOR DEVELOPMENTAL EDUCATION
APPALACHIAN STATE UNIVERSITY
BOONE, NC 28608

Center for Developmental Education
Appalachian State University
Boone, North Carolina 28608

Research in Developmental Education is published five times per academic year.
Editor: Hunter R. Boylan
Managing Editor: Doree N. Pitkin
Consulting Editor: Milton "Bunk" Spann

Manuscripts, news items, and abstracts are accepted by the Editor, RiDE, Center for Developmental Education, Appalachian State University, Boone, NC 28608.

Subscriptions are $9.50 per year. North Carolina residents add 38 cents sales tax; subscribers in foreign countries and $1.50/year shipping and pay by bank draft. Send subscriptions to Managing Editor, RiDE, at the same address or call (704)262-3057.

© 1983

Published by Appalachian State University Volume 3, Issue 1

Research In Developmental Education

Hunter R. Boylan, Editor

Reviewing Learning Styles Inventories

The Canfield and Kolb LSIs

By Hunter R. Boylan

One of the more important trends in developmental education is the increasing use of learning styles assessment. Educators have long known that individual students learn in different ways and at different rates, and a variety of teaching systems have been developed to accomodate individuality in learning. Some of the most notable techniques for individualizing instruction at the college level include Holland and Skinner's "Programmed Instruction" (1961), Keller's "Personalized System of Instruction" (1968) and Postlethwait's "Audio-Tutorial" Instruction" (1969).

All of these systems accomodate individual rates of learning and make some provision to offer different kinds of learning experiences. They did not, however, accomodate individual styles of learning. A major reason for this was that, until fairly recently, there were no instruments available for assessing which students' learning styles might be best served by any given type of instruction.

However, the relatively recent development of several different types of learning styles measures has solved this problem to some extent. Using learning styles assessment, we are now able to determine, with some degree of accuracy, which students learn best from which instructional techniques. This issue of RESEARCH in DEVELOPMENTAL EDUCATION (RiDE) is devoted to a review of two of the more popular learning styles measures: the Canfield Learning Styles Inventory and the Kolb Learning Styles Inventory.

The Canfield LSI

The Canfield LSI was originally developed in 1972 in order to "...measure some of those affective variables that seem to affect learning, and which contribute to satisfactory and effective adjustment to the teaching-learning situation" (Canfield, 1980 p. 1). It should be noted that the Canfield instrument is the only one on the market that emphasizes the affective dimensions of learning as opposed to the cognitive dimensions.

The Canfield LSI measures four dimensions of student preference in learning situtations. These include the conditions of learning, the content of learning, the mode of learning, and student expectations in a learning situation.

Under the category of conditions of learning, the Canfield LSI measures student preferences for:

1. Affiliation - pleasant, friendly, and warm relations with other students or with faculty;
2. Structure - orderly, logical, and well-defined goals, objectives, and study plans;

3. Achievement - independence, self-determined goals and objectives in relation to perceived skills and interests; and

4. Eminence - competition, knowledge of one's own performance in relation to other's, need for control or authority.

Under the category of Content, the Canfield LSI measures student preferences for working with with various sorts of content. These content sub-categories include: numeric, qualitative (working with words or language), inanimate (working with things), and working with people.

Canfield agrees with Gagne's notion of "channel efficiency," the idea that in every individual, some channels of perceiving and processing information are more effective than others (1967). As a result, his instrument also measures students' peferred mode of learning. The categories included under this heading are: listening, reading, iconics (learning through illustrations, movies, slides, graphs, and pictures, etc.), and direct experience.

Finally, the Canfield LSI assesses students expectations of learning - i.e., their anticipated level of performance. The levels of anticipation include outstanding or superior performance, good or above -average performance, average or satisfactory performance, and below-average or unsatisfactory performance.

The instrument measures these categories through 30 items in which students are asked to rank order their preferences among four choices. The structure of the questions requires that students make a "forced choice" in responding. For instance, item 19 in the Canfield LSI requires respondents to rank order to following as a means to learn new material: 1) hearing a lecture, 2) reading a book or text, 3) viewing a movie or slides, or 4) experimenting with a small sample.

Administration of the Canfield LSI generally takes 30 to 45 minutes. The inventory is designed to be self-scoring although the manual suggests that "extra caution should be taken to assure an understanding of how the answers are to be recorded on the separate answer sheet" (Canfield, 1980, p. 9). The inventory package includes the test booklet, answer sheet, and a chart for use in plotting one's learning preferences.

The original version of the Canfield LSI has been criticized because of the reading level of certain items. This was considered to make the test less valid for use with developmental students. A revised version of the Canfield LSI was developed in 1981. The revisions on this form of the Canfield LSI make it much more appropriate for use with developmental students or any other group with poor reading skills.

Reliability

Reliability for the Canfield LSI was established through the use of item analysis, split-half reliability tests, and inter-scale correlations. Each of these tests suggested that the Canfield LSI is a highly reliable instrument. As is shown in TABLE I, the split-half reliabilities were exceptionally high.

Validity

In addition to expert review to establish content validity, a number of studies were undertaken to determine the degree to which students with different majors and backgrounds obtained different score patterns on the Canfield LSI. According to the test manual, "...several statistically significant differences were found between all pairs of the following groups:

1. 52 criminal justice students.
2. 208 business students.
3. 108 education students.
4. 63 physical therapy students.
5. 42 physical therapy faculty.

Additional findings from other studies verifying significant score differences between different groups of students are also reported in the technical manual.

The Kolb LSI

Unlike the Canfield instrument, the Kolb LSI measures cognitive, rather than affective, dimensions of learning styles. The Kolb LSI was derived from Kolb's "theory of experiential learning" (1984).

Kolb conceptualizes the learning process as a series of responses designed to resolve conflicts among four styles of learning. These four styles include:

1. Concrete Experience - the use of sensing and feeling to acquire new information;
2. Reflective Observation - watching and thinking about things in order to learn;
3. Abstract Conceptualization - obtaining information as abstractions and then actively processing new learning; and
4. Active Expeprimentation - doing something with new information or material in order to learn it.

In the Kolb LSI, these four styles of learning are assessed through the use of a questionnaire. The original questionnaire included nine items. As in the Canfield LSI, respondents are asked to make a forced choice among alternatives in each question. Unlike the Canfield LSI which asks respondents for rank-order preferences, the Kolb provides a series of words and asks respondents to indicate which words best describe them. For example, the first item on the Kolb LSI asks respondents to indicate which of the following words are most or least descriptive of their learning style: 1) discriminating, 2) tentative, 3) involved, 4) practical.

The publishers of the Kolb LSI (McBer and Co.) have recently revised the original inventory. The "LSI 1985" includes several changes including expansion in the number of items from nine to twelve and a sentence completion rather than a word choice format. The original version of the Kolb LSI, like the Canfield, was also criticized for the reading level of its test items. The new version is written in much simpler language than the original. The revised version also includes a more simplified scoring format.

Perhaps the most important revisions in the 1985 version are the improved reliability of the instrument and the development of a more representative normative sample. The original instrument was heavily criticized because of its low test-retest reliability. Freedman and Stumpf (1980) state that "Test-retest reliability for the two samples after only three weeks was rather low, (median = .50) suggesting that the LSI is rather volatile" (p. 446).

The 1985 version has also been normed with a much wider sample than the original version. The current edition of the LSI is based on a normative group including various ethnic groups, occupational fields, and income and educational levels. According to the revised technical specifications, the normative group had an "...average education of two years in college" (1985, p. 1).

The Kolb "LSI 85" can be administered to most groups in about half an hour. The instrument is packaged as a booklet which includes a description of the inventory, the inventory questions, instructions for self-scoring, and an explanation of the scores plus a grid for plotting one's learning style.

Validity

There has been considerable debate among psychometrists as to the validity of the Kolb LSI. The items in the inventory were selected to be consistent with Kolb's experiential learning theory. The items were reviewed by Kolb and others to insure content validity - at least according to the theory of experiential learning. Much of the instrument's validity, therefore, is dependent upon the accuracy of Kolb's theory.

As Kolb correctly notes, however, "Learning styles represent preferences for one mode of adaptation over others; but these preferences do not operate to the exclusion of other adaptive modes and will vary from time to time and situation to situation" (1982, p. 4). As a result, the Kolb LSI is simply a straightforward, self-reporting mechanism designed to promote recognition of the complexity of individual approaches to learning and to provide a quick assessment of an individual's preferences at a particular point in time. Since it is not designed to assess fixed traits of individuals, standard techniques for assessing validity may not be applicable to the Kolb LSI.

Reliability

Since Kolb's theory assumes that individual preferences will change and

that the development of a learning style represents an adaptive process, some standard methods of assessing reliability may not be appropriate for the Kolb inventory. Nevertheless, it should be noted that the test-retest reliability of the original version was fairly low (Freedman and Stumpf, 1980).

The new version of the Kolb LSI has been checked for internal reliability using Cronbach's Standardized Scale Alpha. The resulting coefficients are all in the range of .70 to .90. Split-half reliability has also been assessed along with the correlation between the revised inventory and the original inventory. The reliability coefficients all proved to be significant at the .01 level with most of them being in the .80 to .90 range. The results from these assessments are presented in TABLE II.

Summary Comments

In making a decision whether or not to use learning style inventories as part of an assessment process, practitioners should recognize that knowledge of learning styles is at a stage of relative infancy. It can be fairly well established that learning styles do exist.

It can also be established that the learning style inventories discussed here do bear some relationship to subjective reality. In other words, those who take the inventories usually find them to be reflective to some degree of their actual learning preferences.

Unfortunately, no measure of human attributes is completely accurate. Insofar as learning style inventories are concerned, they may be less accurate than other measures simply because our research knowledge of learning styles is still somewhat limited. Nevertheless, they do appear to measure personal learning preferences rather consistently and with some degree of accuracy. They may be imprecise but they do seem to have some validity.

Given this, there are several potentially valid uses for learning styles inventories. They can be used to determine student preferences for learning activities. The Canfield LSI can also be used to determine how students like to have courses organized and delivered, what subjects students prefer to study, and how well students expect to perform academically.

Learning style inventories can be useful, therefore, as part of a pre-assessment program to determine the courses students should take and the types of instruction which seem to have the most potential for success. They can also be used by instructors to determine how their courses may best be organized and delivered in order to maximize student learning.

Learning style inventories are simply tools which can be used well or poorly. They can be useful for a variety of purposes. They should improve our ability to deliver appropriate instruction and to improve the quality of learning among students. They are not, however, thoroughly accurate measures of fixed human traits.

References

Canfield, A.A. LEARNING STYLES INVENTORY MANUAL. Birmingham, MI: Humanics Media, Inc., 1980.

Cross, K.P. BEYOND THE OPEN DOOR. San Francisco: Jossey-Bass, 1971.

Freedman, R.D. and Stumpf, S.A., "Learning Style Theory: Less Than Meets the Eye." ACADEMY OF MANAGEMENT REVIEW, 5, 1980, 445-447.

Gagne, R.M. LEARNING AND INDIVIDUAL DIFFERENCES. Columbus, OH: Merrill, 1967.

Holland, J.G. and Skinner, B.F. THE ANALYSIS OF BEHAVIOR. New York: McGraw-Hill, 1961.

Keller, F.S. "Goodbye Teacher." JOURNAL OF APPLIED BEHAVIOR ANALYSIS. 1968, 1, 79-89.

Kolb, D.A. "Experiential Learning Theory and the Learning Style Inventory: A Reply to Freedman and Stumpf." unpublished paper, 1982.

Kolb, D.A. EXPERIENTIAL LEARNING: EXPERIENCE AS A SOURCE OF LEARNING DEVELOPMENT. Englewood Cliffs, NJ: Prentice-Hall, 1984.

McBer and Company, "Learning-Style Inventory 1985: Technical Specifications." Boston: McBer and Company, 1985.

Postlethwait, S.N., Novak, J. and Murray, H.T. THE AUDIO-TUTORIAL APPROACH TO LEARNING: THROUGH INDEPENDENT STUDY AND INTEGRATED EXPERIENCES (2d Edition). Minneapolis: Burgess, 1969.

TABLE I

Split Half Reliabilities for the Canfield LSI

N = 1397

Scale	First Half Versus Second Half	Odd Numbered Versus Even Numbered Items
Affiliation (peer)	.97	.97
Structure (organization)	.96	.97
Achievement (goal setting)	.97	.97
Eminence (competition)	.98	.98
Affiliation (instructor)	.96	.97
Structure (detail)	.97	.98
Achievement (independence)	.97	.98
Eminence (authority)	.98	.98
Numeric	.98	.98
Qualitative	.98	.99
Inanimate	.98	.98
People	.98	.98
Listening	.98	.97
Reading	.99	.99
Iconic	.98	.98
Direct Experience	.96	.96
Expectancy (A)	.98	.99
Expectancy (B)	.97	.97
Expectancy (C)	.98	.99
Expectancy (D)	.99	.99

Source: LEARNING STYLES INVENTORY MANUAL (Canfield, 1980).

TABLE II

Split Half Reliabilities of the Kolb LSI

N = 268

Category	Split-Half Reliability (Spearman-Brown)	Correlation Between 1976 and 1985 Editions
Concrete Experience	.81	.89
Reflective Observation	.71	.87
Abstract Concepualization	.84	.92
Active Experimentation	.83	.92
Abstract/Concrete	.85	.92
Active/Reflective	.82	.93

Source: LEARNING STYLE INVENTORY FOR 1985: TECHNICAL SPECIFICATIONS, Boston: McBer & Co., 1985.

U.S. Postage
Permit No. 36
PAID
Non-Profit Org.
Boone, N.C.

Center for Developmental Education
Appalachian State University
Boone, North Carolina 28608

Research in Developmental Education is published five times per academic year.
Editor: Hunter R. Boylan
Managing Editor: Doree N. Pitkin
Consulting Editor: Milton "Bunk" Spann

Manuscripts, news items, and abstracts are accepted by the Editor, RiDE, Center for Developmental Education, Appalachian State University, Boone, NC 28608

Subscriptions are $9.50 per year. North Carolina residents add 38 cents sales tax; subscribers in foreign countries add $1.50/year shipping and pay by bank draft. Send subscriptions to Managing Editor, RiDE, at the same address or call (704) 262-3057.

REVIEW OF
RESEARCH
in Developmental Education
Gene Kerstiens, Editor
Published by Appalachian State University
Volume 6, Issue 5, 1989

Assessing Assessment

by Dennis Gabriel

Assessment Perspective

Perhaps because half of all new college students can be classified as academically underprepared, writers have assessed the extent of the problem variously. Abraham (1986) reported 30%, Skinner and Carter (1987), 40%; Lutz (1979), 43%; Bray (1983), half; O'Banion (1988), half; and Haase and Caffrey (1983), 60%. Responding to the reality of those estimates, colleges assess students' skills at matriculation and attempt to place students into "mathemagenically" appropriate reading, writing, and mathematics classes. Woods (1985) reported that an American College Testing and American Association of Community and Junior Colleges study found over 90% of two-year colleges offer assessment. Although Lederman, Ribaudo, and Ryzewic's, (1983) national study found that 97% of all colleges offer assessment, their study, Skinner and Carter's (1987) Texas study, and Rounds and Anderson's (1984) California study confirmed that fewer than two-thirds of all colleges require assessment, as shown in Table 1.

Table 1

Percent of Colleges with Mandatory Testing*

California		Texas**	National
Reading	32	2	69 (91% of the 76% that offer basic reading)
Writing	56	1	65 (85% of the 77% that offer basic writing)
Math	25	2	64 (86% of the 74% that offer basic math)

*Colleges may not assess all students: The student may already have a degree, be a transfer student, or have taken an admissions test in lieu of placement (Skinner & Carter, 1987).

**For the Texas study, colleges testing 100% of new students were defined as having mandatory testing.

Understandably, there is uncertainty concerning the scope and scale of assessment practices. For instance, to say that 90% of all colleges offer assessment is misleading since at least one-third of all colleges do not have mandatory assessment. For example, in Texas where the percentage of assessed new students ranged from 8 to 100%, the typical school assessed 43% of new students in reading, 40% in writing, and 45% in math. Only three colleges did not assess the basic skills, yet fewer than two percent of Texas' two-year schools mandate assessment (Skinner & Carter, 1987). Our knowledge of practices in assessment testing also suffers from other complications and distractions such as myriad cutoff scores on tests and sparse empirical justification for assessment tests.

Assessment Instruments and Strategies

As shown in Table 2, two state, one area, and a national study examined the diversity of assessment tests: Rounds and Anderson's (1984) survey of California community colleges; a Texas study of two-year colleges (Skinner & Carter, 1987); Southern Regional Educational Board's (SREB) survey (Abraham, 1986) which yielded 100 combinations of tests; and a national study (Lederman, Ribaudo, & Ryzewic, 1983) of 1,269 colleges.

Discussion of surveys

Expectedly, the surveys show that the ND is the most common reading test; beyond that, the surveys demonstrate state and regional interests. For instance, the Coop-Reading, which was used in 7 out of 99 California colleges, was not mentioned in the SREB study. The SAT, on the other hand, was used once for reading assessment among the 99 California colleges.

The research makes a weak case for reading tests' use and ability to predict success in classes. First, the ND was not without criticism. A timed test such as the ND is not appropriate for developmental students (Clary, 1973; Kerstiens, 1986a, 1986b). Due to ND cutoff scores as low as grade 7, Abraham (1986) found *college level* to be meaningless; the lowest cutoff level allows 99% of all students to enter college-level classes. Second, scores do not correlate well with students' success. Santa Rosa

Table 2

Reported Frequency* of Assessment Tests**

California		Texas		Southern Area		National***	
Reading							
N	**Test**	**N**	**Test**	**N**	**Test**	**N**	**Test**
18	ND	24	ND	121	ND	236	ND
9	CGP	13	ASSET	89	ACT-Misc.	152	Local test
7	Coop-Reading	7	DTLS	37	DTLS	81	ACT
6	Davis	4	SDRT	35	SAT-Verbal	55	SAT-Verbal
5	Local test	4	DAT	29	ASSET	47	CGP
14	Other	7	Other	62	Other	106	Other
Writing/English							
N	**Test**	**N**	**Test**	**N**	**Test**	**N**	**Test**
64	Essay	13	ASSET	108	ACT-Misc.	365	Local test
18	Local test	9	DTLS	66	Local test	127	ACT
10	Coop-Eng./Rd.	6	ND	57	Essay	98	SAT-Verbal
9	CGP	3	TSWE	53	TSWE	96	Essay
8	ND	3	WEEP	28	SAT-Verbal	62	TSWE
32	Other	9	Other	91	Other	84	Other
Mathematics							
N	**Test**	**N**	**Test**	**N**	**Test**	**N**	**Test**
26	Local test	14	DTMS	118	Local test	393	Local test
10	SCAT-Math	13	ASSET	97	ACT-Misc.	115	ACT
9	CGP	5	MAA	85	DTMS-Misc.	78	SAT-Math
3	Coop-Math	4	DAT	47	SAT-Math	40	State test
6	Other	1	Other	29	State test	40	CGP
	NA		NA	40	Other		NA

Junior College (1984) reported that reading tests did not predict success in any course. Yamagishi and Gillmore (1980) studied the ND and a combination of test scores and writing samples upon academic success and found no effective predictor of academic success. They concluded the ND may lack predictive validity.

Although local writing tests were widely used, the studies again demonstrate state and regional interests. For example, 10 California colleges used the Coop-English or Reading, but the Coop was not listed among the 30 writing tests in the SREB study.

The research shows that colleges have problems with writing assessment. First, colleges may use questionable assessment devices. Six Texas and eight California colleges used the ND—a reading test—to place students in writing classes; the Davis and Coop-Reading were also used to place students in writing classes in California. Second, writing samples alone are not adequate for identifying basic-level students (Gordon, 1987). When two

Legend

American College Testing (ACT)
Assessment of Skills for Successful Entry and Transfer (ASSET)
Comparative Guidance and Placement Battery (MAPS—CGP)
Cooperative (Coop)
Cooperative School College Ability Test (SCAT)
Descriptive Test of Language Skills (MAPS—DTLS)
Descriptive Test of Math Skills (MAPS—DTMS)
Differential Aptitude Tests (DAT)
Mathematical Association of America (MAA)
Multiple Assessment Programs and Services (MAPS)
Nelson Denny (ND)
Scholastic Aptitude Test (SAT)
Stanford Diagnostic Reading Test (SDRT)
Test of Standard Written English (TSWE)
Written English Expression Placement Test (WEEP)

*As Gordon (1987), Guerrero and Robinson (1986), and Olson and Martin (1980) noted, a single assessment is not adequate for placement; therefore, some colleges give tests in combination. Thus, percentages can be confusing.

**The fact that a college offers a test does not mean that the test is mandatory. Also, mandatory placement does not always follow mandatory testing.

***For the national study, only those tests used by at least 20 colleges were included in the statistics.

instructors read the same papers, Guerrero and Robinson (1986) reported that instructors disagreed on course placement 43% of the time. Even when writing samples are used with objective scores, Olson and Martin (1980) found only 39% of the students received the same placement recommendation. Third, there is a lack of correlation between writing samples and achievement (Alexander & Swartz, 1982).

Local math tests were widely used, but, again, state and regional differences abounded. Widely used in the state and SREB studies, ASSET, SCAT, and DTMS were each used in fewer than three percent of the colleges in the national study.

Selecting cutoff points and proving that math testing works presented challenges to colleges. Abraham (1986) cited evidence that cutoff scores ranged from 1 to 18 on the DTMS; at the lower level, all but 14% of the students could be placed in college-level classes. There is no correlation between placement scores and final mathematics grades (Sworder, 1986).

Even widely used and commonly cited admissions tests are questionable placement tools. Morante (1987) cited the New Jersey Basic Skills Council's findings that many students who had scored above-average on the SAT still were not ready for college-level classes. Grulick (1986) reported mixed findings after a review of the SAT as an assessment tool and argued for local tests.

High-school grades

In nontest evaluations, high-school grades are used in 43% of the cases for reading, 46% for writing, and 56% for mathematics (Lederman, Ribaudo, & Ryzewic, 1983). The use of high-school grades for placement is not appropriate (Morante, 1987) since grades provide an inflated view of students' abilities (Roueche, Baker, & Roueche, 1986), and the proficiency required to finish three years of high-school English or mathematics is considerably lower than the level expected of most college freshmen (Edge, 1979).

Summary

The research demonstrates regional test selection differences but little evidence to prove the effectiveness of assessment testing or agreement as to what constitutes college level. With cutoff scores as low as seventh grade, college students resemble Lake Wobegon's children: they are all above average. Perhaps the best assessment approach is a combination of several measures of each basic skill plus a professional analysis of the results, and on this point, the literature is weak.

Making Assessment Work

Nevertheless, several inferences about successful assessment can be drawn. One characteristic of successful assessment is a plan to meet the local assessment challenge. Lederman, Ribaudo, & Ryzewic (1985) proffered entry-level testing, prescriptions, and exit testing. Bray (1983) explained California's Learning Assessment Retention Consortium's plans to develop assessment philosophies, establish goals and objectives, set up assessment centers, select tests, and assess students.

Second, mandatory testing and placement are essential, but as noted in Table 1, at best only two-thirds of college students face mandatory assessment and placement. Lum and Alfred (1987) reported that students in compulsory programs were more likely to be persisters and perform well on long-term achievement measures than students in voluntary programs. For 1,300 students, Richards (1986) found that 73% followed placement advice and succeeded in the recommended classes; 15% followed advice and did not succeed; 6% neither followed advice nor succeeded; and 6% did not follow advice but succeeded. Roueche, Baker, & Roueche (1986) reported that the majority of colleges favored mandatory assessment but were not strongly in favor of mandatory placement; they conclude that assessment is futile without mandatory placement.

Third, in addition to testing basic skills, colleges should consider other instruments that survey students' study habits and academic confidence. Accordingly, Bliss and Mueller (1987) discussed a promising assessment device, the Study Behavior Inventory, which measures short-term and long-term study behaviors as well as academic self-perception. Scores on this instrument correlate with GPA at .79.

Fourth, good programs use technology to improve assessment and research. Rounds, Kanter, and Blumin (1987) note ACT is designing new components for computer-adaptive testing. ETS' computer-adaptive test (Computerized Placement Test) has reading, sentence skills, mathematical reasoning, and algebra components. In the technologically ideal assessment, correct answers trigger more difficult questions, and incorrect answers trigger easier questions. Thus, frustration, instructional, and independent levels can quickly be determined. With the new technology, Wainer (1983) noted that students can be tested at any time, results are instantaneous, test security is improved, students can work at their own pace, there are no problems with answer sheets, but there is a tendency to rely on one instrument for placement. At the University of California, Irvine (Shoemaker, St. John, & Lewis, 1986), computers report means and percentages of students placed in courses and reliability and item analyses of placement tests. UCI tests in chemistry, mathematics, reading, writing, and ESL.

Proof that Assessment Works

The use of end-of-quarter grades to demonstrate assessment effectiveness requires more thought. Rasor and Powell (1984) reported no correlation between test scores and final grades for three out of four classes.

Sworder (1986) found no correlation between placement scores and final math and English grades. Palmer (1987) concluded that most studies found only a low correlation between assessment tests and grades. The San Diego Community College District study (1983) found a significant relationship between test scores and grades in English 101 only. However, Morante (1987) noted that the correlation between assessment scores and grades should be close to *zero* in a good remedial program.

The best evidence in favor of mandatory assessment and placement comes from Roueche and Baker (1986) who noted that students at Miami-Dade who took suggested developmental classes had a nine times better chance of graduating than students who declined to take developmental classes.

Conclusions

Because of myriad cut-off points on placement tests, *college level* is ambiguous. Nevertheless, half of all college students require developmental work. Although assessment testing is widespread, there is no reason to conclude that as many as two-thirds of all colleges have mandatory testing and mandatory placement for reading, writing, and mathematics. The use of a single placement device, such as high-school grades, writing samples, or admissions scores, provides an inaccurate view of students' abilities. A good approach to assessment involves a planned program, mandatory tests for each of the basic skills, mandatory placement, use of current technology, program assessment, a survey of study habits and attitudes, and dissemination of testing information. Proof that tests or writing samples predict academic success remains sparse. Empirical evidence that assessment works is not to be found in the correlation between placement scores and final grades (the correlation is close to zero in many cases). However, if mandatory assessment and mandatory placement can improve a student's chances of graduation by a factor of nine—as Roueche and Baker's (1986) review of Miami-Dade experiences suggests—then mandatory assessment and mandatory placement are essential.

Works Cited

Abraham, A.A. (1986). *College-level study: What is it? Variations in college placement tests and standards in the SREB states* (Report No. 22). Atlanta, GA: Southern Regional Educational Board. (Eric Document Reproduction Service No. ED 277 715)

Alexander, J., & Swartz, F. (1982, October). *The dynamics of computer-assisted writing sample measurement at Ferris State College.* Paper presented at the meeting of the Michigan Council of Teachers of English, East Lansing, MI. (ERIC Document Reproduction Service No. ED 233 344)

Bliss, L.B., & Mueller, R. (1987). Assessing study behaviors of college students: Findings from a new instrument. *Journal of Developmental Education, 11*(2), 14-18.

Bray, D. (1983). *Assessment/placement: A design for student success and institutional planning for the 1980's Management Report, 1983-4.* Sacramento: CA: Association of California Community Colleges. (ERIC Document Reproduction Service No. ED 230 234)

Clary, L. (1973). *The Nelson Denny Test in college reading programs.* Augusta, SC: Augusta College. (ERIC Document Reproduction Service No. ED 177 488)

Edge, D. (1979). *Report of the New Jersey Basic Skills Council.* Trenton, NJ: New Jersey State Dept. of Higher Education. (ERIC Document Reproduction Service No. ED 185 098)

Gordon, B.L. (1987). Another look: Standardized tests for placement in college composition courses. *Writing Program Administration Journal, 3*(29), 29-38.

Grulick, L.E. (1986). *Evaluation of admission and placement testing at Florence-Darlington Technical College.* Columbia, SC: University of South Carolina. (ERIC Document Reproduction Service No. ED 277 431)

Guerrero, B., & Robinson, R. (1986). *Holistic evaluation of writing samples for placement in post-secondary English composition courses.* Honolulu, HI: Hawaii Educational Research Association. (ERIC Document Reproduction Service No. ED 270 140)

Haase, M., & Caffrey, P. (1983). *Assessment and placement at Sacramento City College.* Sacramento, CA: Sacramento City College. (ERIC Document Reproduction Service No. ED 230 258)

Kerstiens, G. (1986a). A testimonial on timed testing: Developmental students and reading comprehension tests. In M. Douglass (Ed.), *Fiftieth Yearbook of the Claremont Reading Conference* (pp. 261-268). Claremont, CA: Claremont Graduate School.

Kerstiens, G. (1986b). *Time-critical reading comprehension tests and developmental students.* Paper delivered at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service no. ED 278 700)

Lederman, M., Ribaudo, M., & Ryzewic, S. (1983). *Assessment and improvement of the academic skills of entering freshmen: A national survey* (Research Monograph Series Report No. 5). New York: City University of New York. (ERIC Document Reproduction Service No. ED 238 973)

Lederman, M., Ribaudo, M., & Ryzewic, S. (1985). Basic skills of entering college freshmen: A national survey. *Journal of Developmental Education, 9*(1), 10-13.

Lum, G., & Alfred, R. (1987). *Remedial program policies, student demographic characteristics and performance outcomes in community colleges.* San Diego, CA: Association for the study of Higher Education. (ERIC Document Reproduction Service No. ED 281 452)

Lutz, W.D. (1979). *Statewide testing in New Jersey.* San Francisco, CA: Modern Language Association. (ERIC Document Reproduction Service No. ED 181 485)

Morante, E.A. (1987). A primer on placement testing. In D. Bray & M. Belcher (Eds.), *New directions for community colleges: Issues in student assessment* (pp. 55-63). San Francisco, CA: Jossey Bass.

O'Banion, T. (1988). Leadership through innovation: Improving the record on underprepared students. *National Association for Developmental Education Newsletter, 12*(1), 1-6.

Olson, M., & Martin, D. (1980). *Assessment of entering student writing skill in the community college.* Boston, MA: American Educational Research Association. (ERIC Document Reproduction Service No. ED 235 845)

Palmer, J. (1987). Sources and information: Student assessment at community colleges. In D. Bray & M. Belcher (Eds.), *New directions for community colleges: Issues in student assessment* (pp. 103-112). San Francisco, CA: Jossey Bass.

Rasor, R., & Powell, T. (1984). *Predicting English writing course success with the Vocabulary and Usage subtests of the Descriptive Tests of Language Skills of the College Board.* Sacramento, CA: American River College. (ERIC Document Reproduction Service No. ED 218 341)

Richards, W. (1986). *The effectiveness of new-student basic skills assessment in Colorado community colleges.* Denver, CO: Colorado State Board for Community Colleges and Occupational Education, Assessment and Basic Skills Committee. (ERIC Document Reproduction Service No. ED 275 351)

Roueche, J., & Baker, G. (1986). The success connection: Examining the fruits of excellence. *Community, Technical, and Junior College Journal, 56*(5), 47-56.

Roueche, J.E., Baker, G., & Roueche, S. (1986). Access with excellence. In G.H. Vogel (Ed.), *New directions for community colleges* (pp. 29-40). San Francisco: Jossey-Bass.

Rounds, J.C., & Anderson, D. (1984). *Tests in use in California Community Colleges: Standardized tests most used for placement in English, reading, ESL and math.* Marysville, CA: Yuba College. (ERIC Document Reproduction Service No. ED 250 037)

Rounds, J., Kanter, M., & Blumin, M. (1987). Technology and testing: What is around the corner? In D. Bray & M. Belcher (Eds.), *New directions for community colleges: Issues in student assessment* (pp. 83-93). San Francisco, CA: Jossey Bass.

San Diego Community College District. (1983). *The English placement test: A correlation analysis.* San Diego, CA: San Diego Community College District. (ERIC Document Reproduction Service No. ED 239 669)

Santa Rosa Junior College. (1984). *DRT/ASSET/final grade study: Fund for instructional improvement final report, 1983-84.* Santa Rosa, CA: Santa Rosa Junior College. (ERIC Document Reproduction Service No. ED 253 272)

Skinner, E., & Carter, S. (1987). *Texas association of junior and community college instructional administrators' remedial education study.* Tempe, AZ: Arizona State University. (ERIC Document Reproduction Service No. ED 291 416)

Shoemaker, J., St. John, E., & Lewis, R. (1986). *Using microcomputer software to score placement tests—an example from the University of California, Irvine.* San Francisco, CA: American Educational Research Council. (ERIC Document Reproduction Service No. ED 270 472)

Sworder, S. (1986). *Design and application of evaluation procedures for the Saddleback Community College District's Assessment and Placement Program.* (Practicum) Ft. Lauderdale: Nova University. (ERIC Document Reproduction Service No. ED 272 246)

Wainer, H. (1983). On item response theory and computerized adaptive tests. *Journal of College Admissions, 28*(4), 9-16.

Woods, J.C. (1985). *Status of testing practices at two-year postsecondary institutions.* Iowa City, IA: ACT Publications. (ERIC Document Reproduction Service No. ED 264 907)

Yamagishi, M., & Gillmore, G. (1980). *The relationship between Nelson Denny test scores and academic performance of educational opportunity program students.* Seattle, WA: Educational Assessment Center. (ERIC Document Reproduction Service No. ED 197 656)

Acknowledgement

Dennis Gabriel, professor of English, Jianghan University, Wuhan, Hubei, People's Republic of China.

TELEMENTORING RESOURCES

CONTINUING PROFESSIONAL DEVELOPMENT

Videotapes, study guides, and annotated bibliographies for:

- in-service training;
- professional development;
- faculty orientation

in developmental and learning assistance programs

For Additional Information Contact:

Wilma Brinegar
Director of Marketing
National Center for Developmental Education
Appalachian State University
Boone, NC 28608
(704) 262-3058

U.S. Postage
Permit No. 36
PAID
Non-Profit Org.
Boone, N.C.

National Center for Developmental Education
Reich College of Education
Appalachian State University
Boone, North Carolina 28608

Review of Research in Developmental Education is published five times per academic year.
Editor: Gene Kerstiens
Managing Editor:
Barbara Calderwood

Manuscripts, news items, and abstracts are accepted by the Editor, RRIDE, National Center for Developmental Education, Appalachian State University, Boone, NC 28608

Subscriptions are $9.50 per year. North Carolina residents add 38 cents sales tax; subscribers in foreign countries add $1.50/year shipping and pay by bank draft. Send subscription to Managing Editor, RRIDE, at the same address or call (704)262-3057.

ERIC Clearinghouse for Junior Colleges
MAR 20 1992